

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 9, 2003, 08:33:19 ; Search time 819 seconds
(without alignments)
11317.735 Million cell updates/sec

Title: US-09-601-965C-1
Perfect score: 4116
Sequence: 1 aaaaaaaaaagctccccc.....cattatgcatctcattcgtc 4116

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq_101002.*
1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4116	100.0	4116	20	AAx86690 DNA encoding a TAF
2	91	2.2	3399	17	AAT05868 Chicken leucocytos
3	88.6	2.2	1998	21	AAA70212 Plasmodium falcipar
4	86.6	2.1	3579	21	AAA70099 Plasmodium falcipar
5	84	2.0	4056	21	AAA70225 Plasmodium falcipar
6	82.2	2.0	1686	16	AA087587 DNA encoding leuco
7	79.8	1.9	1080	24	AB199537 Mouse ischaemic co
8	79.2	1.9	49999	20	AA23895 Murine LOB8 homolo
9	78.4	1.9	400	23	AA575460 DNA encoding novel

10	78.4	1.9	963	23	AA568580
11	78.2	1.9	801	23	AA590729
12	77.6	1.9	1083	23	AA516745
13	77	1.9	372	23	AA575452
14	77	1.9	1072	23	AA590738
15	76.8	1.9	700	20	AA588907
16	76.8	1.9	1391	21	AA257866
17	76.8	1.9	2554	21	ABA92314
18	76	1.8	5994	21	AAA70222
19	76	1.8	6621	21	AAA70188
20	75.6	1.8	4167	22	AA589234
21	75.6	1.8	4167	22	AA530018
22	75.6	1.8	4167	22	AA105022
23	75.6	1.8	4167	22	AA107543
24	75.6	1.8	4167	22	AA528711
25	75.6	1.8	4167	22	AAK68040
26	75.6	1.8	4167	22	AAK9382
27	75.6	1.8	4167	23	ABL97915
28	75.6	1.8	4167	23	ABL3697
29	75.2	1.8	6668	24	AA032658
30	75.2	1.8	507	13	AAH84231
31	75.2	1.8	699	22	AAH84230
32	75.2	1.8	717	22	AAH84229
33	75.2	1.8	717	22	AAH84228
34	75.2	1.8	819	22	AAH84227
35	75.2	1.8	1669	22	AAH84226
36	75.2	1.8	2427	23	AA569554
37	75.2	1.8	2427	23	AA588862
38	75.2	1.8	14006	24	ABL33958
39	75.2	1.8	37996	23	ABL07876
40	75	1.8	510	23	AA569539
41	74.8	1.8	510	23	AA571141
42	74.8	1.8	510	23	AA590687
43	74.8	1.8	1527	21	AAA70121
44	74.2	1.8	3738	21	AAA70178
45	74.2	1.8			

ALIGNMENTS

RESULT 1	AAx86690	standard; DNA; 4116 BP.
ID	AAx86690;	
AC	AAx86690;	
XX		
DT	14-OCT-1999	(first entry)
XX		
DE	DNA encoding a TAF-145 protein of Candida albicans.	
XX		
KM	TAF-145 protein associated factor; TAF-145; gene transcription;	
KW	fungus gene transcription; fungal infection; ss.	
XX		
OS	Candida albicans.	
XX		
FM	Key	Location/Qualifiers
FT	CDS	330..3812
FT		/*tag= a
XX		/transl_except= (pos: 3267..3269, aa: Ser)
PN	WO9940199-A2.	
PD	12-AUG-1999.	
XX		
PF	08-FEB-1999;	99WO-US02940.
XX		
PR	09-FEB-1998;	98US-0024100.
XX		
PA	(SCRI-) SCRIPTGEN PHARM INC.	
XX		
PI	Long F, Thompson CM, Wobbe RC;	
XX		

DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
EST clone IN200.
Protein regulating
Human cylicin 30 p
Plasmodium falcipa
Plasmodium falcipa
Human ovarian and
Genomic sequence #
Human lung antigen
Human reproductive
Human reproductive
Genomic sequence #
Human immune/heama
Human digestive sy
Human testicular a
Human immune syste
LEMMI 10 CDNA. Ly
Human cell death p
Human cell death p
Human cell death p
Human cell death p
Human cell death p
Human cell death p
DNA encoding novel
DNA encoding novel
Human immune syste
Drosophila melanog
DNA encoding novel
DNA encoding novel
DNA encoding novel
Plasmodium falcipa
Plasmodium falcipa

606741100

1801 CATGTGAGCTTTTCTAAATTGAAGTTCGGAAAGCGAAAAAACAAGGGAAATCTT 1860
1801 CATGTGAGCTTTTCTAAATTGAAGTTCGGAAAGCGAAAAAACAAGGGAAATCTT 1860
1861 TGCAACAGATTTTTCCTAAACAGTGAAGTTCAGTTCGTGATACGTGTAATTAATTG 1920
1861 TGCAACAGATTTTTCCTAAACAGTGAAGTTCAGTTCGTGATACGTGTAATTAATTG 1920
1861 TGCAACAGATTTTTCCTAAACAGTGAAGTTCAGTTCGTGATACGTGTAATTAATTG 1920
1921 CTTAGAGTATTCGCAACAGTATCCAACTTTTATCGAATTTTCGGATGGGGTGAAT 1980
1921 CTTAGAGTATTCGCAACAGTATCCAACTTTTATCGAATTTTCGGATGGGGTGAAT 1980
1981 TGATTAATCTATATCTGAAGAAAGAACCAACGACCTTCACTGTCACAGGCTCAATTTG 2040
1981 TGATTAATCTATATCTGAAGAAAGAACCAACGACCTTCACTGTCACAGGCTCAATTTG 2040
2041 GGGAACTCATATTTTGGGGGTGGAGATAGATCCCATTTTGGAAATTCGGTGAAGTTG 2100
2041 GGGAACTCATATTTTGGGGGTGGAGATAGATCCCATTTTGGAAATTTGGTGAAGTTG 2100
2101 CTCTGAGATTTTGTCCCACTGTAATTAATATGTAAGACCAATTTTCAAGC 2160
2101 CTCTGAGATTTTGTCCCACTGTAATTAATATGTAAGACCAATTTTCAAGC 2160
2161 ATGACAAACAACCACTGATTTTCTATTGTTAATCTCAAGGGCTGGATCGCAACAA 2220
2161 ATGACAAACAACCACTGATTTTCTATTGTTAATCTCAAGGGCTGGATCGCAACAA 2220
2221 AATTTATTTTGGGAGGATCAATTTCACTTGTGTGTAACAATTTCCAGTTGAG 2280
2221 AATTTATTTTGGGAGGATCAATTTCACTTGTGTGTAACAATTTCCAGTTGAG 2280
2281 TTCCAGCTCTCTACTCGAGAAAAGTGAACAATATCTCAAAAAATAGGTTGAAAATGTAAG 2340
2281 TTCCAGCTCTCTACTCGAGAAAAGTGAACAATATCTCAAAAAATAGGTTGAAAATGTAAG 2340
2341 TTTTCAAGTGAATGTAATAGCTTAAGAGTGCACGATTTCTGTGAAGATGTTTCAAAAC 2400
2341 TTTTCAAGTGAATGTAATAGCTTAAGAGTGCACGATTTCTGTGAAGATGTTTCAAAAC 2400
2401 ATTTTCTGAGCATAGCGATATGCAAAATCGACAAAGATTGAAAAGATTTATGAAATACC 2460
2401 ATTTTCTGAGCATAGCGATATGCAAAATCGACAAAGATTGAAAAGATTTATGAAATACC 2460
2461 AAAAGCAAGGTGAGAACCAAGGATTTGAAAGTAAGAGATTTAAACGATGTAATTCAG 2520
2461 AAAAGCAAGGTGAGAACCAAGGATTTGAAAGTAAGAGATTTAAACGATGTAATTCAG 2520
2521 GAGAAGAGAAATTCGAATATGATTAACACAGAGATTTCTTCAATGATGACACAATGC 2580
2521 GAGAAGAGAAATTCGAATATGATTAACACAGAGATTTCTTCAATGATGACACAATGC 2580
2521 GAGAAGAGAAATTCGAATATGATTAACACAGAGATTTCTTCAATGATGACACAATGC 2580
2581 AATTTGACACCAAGTACTAGATGATATATGTTATTCGGTGAACAGACAGCAAG 2640
2581 AATTTGACACCAAGTACTAGATGATATATGTTATTCGGTGAACAGACAGCAAG 2640
2641 AATCATCTAGAAAGTGAAGAAAGTGAATTAAGAGAGATTCATTTGGGATGATGCAAGAA 2700
2641 AATCATCTAGAAAGTGAAGAAAGTGAATTAAGAGAGATTCATTTGGGATGATGCAAGAA 2700
2701 ATGAGATGATATTAATTAAGATTAAGAGAAAGGTTGGAAGAAAGAAAGACAGAAA 2760
2701 ATGAGATGATATTAATTAAGATTAAGAGAAAGGTTGGAAGAAAGAAAGACAGAAA 2760
2761 GAGAAGAGAAAGGAGAAAGATTAAGAGAAAGACAGAGCAAGGAAAGACAAAACCG 2820
2761 GAGAAGAGAAAGGAGAAAGATTAAGAGAAAGACAGAGCAAGGAAAGACAAAACCG 2820
2821 AGAAGAGAAATCAAAAGAAATCAAAAGAGACAGACTGAAATTTGATGTAAGAGAAAT 2880
2821 AGAAGAGAAATCAAAAGAAATCAAAAGAGACAGACTGAAATTTGATGTAAGAGAAAT 2880
2881 TGGAACCATGGAATTTATCGAAGAAATTTTGTATCGCAATCAGACCAACAAATGCTAC 2940

2881 TGCAACCATGGAATTTATCGAAGAAATTTTGTATGCGCAATAGACCAAGACAAATCTAC 2940
2941 AATTTGAATGTTGAAGGTGATCCAACTGGAATTTGGTATAGGATTTTCATGTTGAGGGCTA 3000
2941 AATTTGAATGTTGAAGGTGATCCAACTGGAATTTGGTATAGGATTTTCATGTTGAGGGCTA 3000
3001 CACAAAAGAACCCATTCAAACCGTTATTTACCCACCAACGAAAATGTCCTTAAAGTA 3060
3001 CACAAAAGAACCCATTCAAACCGTTATTTACCCACCAACGAAAATGTCCTTAAAGTA 3060
3061 ATCTGAGCCCATTAATCAAAAGTTGTAAGAACAGATTAAGAAATATGTAATCTTC 3120
3061 ATCTGAGCCCATTAATCAAAAGTTGTAAGAACAGATTAAGAAATATGTAATCTTC 3120
3121 AAGAAAGCTCTTTAGTTGATCATGGGAAAGAACTGAATCAAGTTGCAACAGATCTATA 3180
3121 AAGAAAGCTCTTTAGTTGATCATGGGAAAGAACTGAATCAAGTTGCAACAGATCTATA 3180
3181 ATGAGTACCCGCGAGAGATCATGAATTTGAAAAAACAACCTTGAAACAAGACCAAC 3240
3181 ATGAGTACCCGCGAGAGATCATGAATTTGTAAGAAACAACCTTGAAACAAGACCAAC 3240
3241 AGGTACAAACAACAACAAGATCTCTGCTTCAAGCTGATCAGCAACAACAAC 3300
3241 AGGTACAAACAACAACAAGATCTCTGCTTCAAGCTGATCAGCAACAACAAC 3300
3301 AACAAACAATCGAGTTTGAAGATTACTAGAAAGTCCGAGATGAATGAATATGTGC 3360
3301 AACAAACAATCGAGTTTGAAGATTACTAGAAAGTCCGAGATGAATGAATATGTGC 3360
3301 AACAAACAATCGAGTTTGAAGATTACTAGAAAGTCCGAGATGAATGAATATGTGC 3360
3361 ATGAAAGGTTGAATTAATCATATCATGATCCAGATTAATAGAGCATATGTAAGCGTAAGA 3420
3361 ATGAAAGGTTGAATTAATCATATCATGATCCAGATTAATAGAGCATATGTAAGCGTAAGA 3420
3421 AACAAATGGAAGTGAATTTGAAGATGCTGAATGTTGATGAATATTAACCTACTAACG 3480
3421 AACAAATGGAAGTGAATTTGAAGATGCTGAATGTTGATGAATATTAACCTACTAACG 3480
3481 ACAAGAGTTTAAACAAAATCCCTGCTGAAGATTTGAAGAAAGTTGGCTAATTTGGAGA 3540
3481 ACAAGAGTTTAAACAAAATCCCTGCTGAAGATTTGAAGAAAGTTGGCTAATTTGGAGA 3540
3481 ACAAGAGTTTAAACAAAATCCCTGCTGAAGATTTGAAGAAAGTTGGCTAATTTGGAGA 3540
3541 AACGGGCAAAAACAAGCGGGGCCAAAAAACAACAAGGATCTGATCCATGCAAGCTGCG 3600
3541 AACGGGCAAAAACAAGCGGGGCCAAAAAACAACAAGGATCTGATCCATGCAAGCTGCG 3600
3601 AACGGGCAAAAACAAGCGGGGCCAAAAAACAACAAGGATCTGATCCATGCAAGCTGCG 3600
3601 AACGGGCAAAAACAAGCGGGGCCAAAAAACAACAAGGATCTGATCCATGCAAGCTGCG 3600
3661 TTGGTGTAAAGGTTATTTGTAAGGGGAAAGTCAACCTGCTGTTTAAAAATTTGGAG 3720
3661 TTGGTGTAAAGGTTATTTGTAAGGGGAAAGTCAACCTGCTGTTTAAAAATTTGGAG 3720
3721 CTTATGACACATTCGTACTAATGCAAAATGTCCTTATATAGAAATTTGGCTTGA 3780
3721 CTTATGACACATTCGTACTAATGCAAAATGTCCTTATATAGAAATTTGGCTTGA 3780
3781 TCGATGATATTCAGCGGCGAGTTGTTGTTGACACAGCAATGTTAGTGAAGTGAAGTAT 3840
3781 TCGATGATATTCAGCGGCGAGTTGTTGTTGTTGACACAGCAATGTTAGTGAAGTGAAGTAT 3840
3841 TGGAGAAACAACAACATCTACCGCAGTAATCTGTAATACACAGGATTCAGAGAACAGAA 3900
3841 TGGAGAAACAACAACATCTACCGCAGTAATCTGTAATACACAGGATTCAGAGAACAGAA 3900
3841 TGGAGAAACAACAACATCTACCGCAGTAATCTGTAATACACAGGATTCAGAGAACAGAA 3900
3901 TCTGGCTGAAGGCTGACCAATGCAATCAAAAATAGATATGAACCCAGAGATAGAGTTA 3960
3901 TCTGGCTGAAGGCTGACCAATGCAATCAAAAATAGATATGAACCCAGAGATAGAGTTA 3960
3961 CTGAGTTGCAAAATTTAGAGTGAATTCGCAATGCAATTTTGTCTATCAAAAACGACATGA 4020

GenCore version 5.1.4 ps.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 08:34:49 ; Search time 10434 Seconds

(without alignments)
11480.463 Million cell updates/sec

Title: US-09-601-965C-1

Perfect score: 4116

Sequence: 1 aaaaaaaaaaacgtccccc.....tattatgcattcttattgtc 4116

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_ets:*

12: gb_ey:*

13: gb_un:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_scs:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rnd:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_ey:*

39: em_hcg_hum:*

40: em_hcg_mus:*

41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1767.2	42.9	1794	6 AX489468	AX489468 Sequence
2	296.8	7.2	3950	8 SCYGR274C	Z73059 S.cerevisia
3	296.8	7.2	7077	8 SCU14954	U14954 Saccharomyc
4	296.8	7.2	8061	8 SCYCCG1	X84098 S.cerevisia
5	275.8	6.7	36162	8 SPAC2G11	Z54354 S.pombe chr
6	125.6	3.1	811	11 CNS06EWE	AL395652 T3 end of
7	114.2	2.8	7218	6 166494	166494 Sequence 14
8	108.4	2.6	110000	2 PPMAL4P1_1	Continuation (2 of
9	105.4	2.6	115758	9 AC104634	AC104634 Homo sapi
10	103.4	2.5	43993	2 AC116965	AC116965 Dictyoste
11	99.2	2.4	62352	2 AC116990	AC116990 Dictyoste
12	99	2.4	224236	2 AL806518	AL806518 Mus muscu
13	98.2	2.4	309587	2 AL713990	AL713990 Mus muscu
14	97.4	2.4	59004	2 AC115680	AC115680 Dictyoste
15	96.6	2.3	219751	2 AL732595	AL732595 Mus muscu
16	96.6	2.3	259744	2 AL844181	AL844181 Mus muscu
17	96.2	2.3	230749	2 AC124415	AC124415 Mus muscu
18	95.8	2.3	155204	2 AC007926	AC007926 Trypanoso
19	95.2	2.3	155682	2 AC131189	AC131189 Mus muscu
20	94.8	2.3	160759	2 AC117082	AC117082 Dictyoste
21	94.4	2.3	111554	8 AP002460	AP002460 Arabidops
22	94	2.3	59869	2 AL844525	AL844525 Mus muscu
23	93.8	2.3	199285	2 AL844592	AL844592 Mus muscu
24	93.4	2.3	64693	2 AC103396	AC103396 Mus muscu
25	92.8	2.3	163860	2 AC021628	AC021628 Mus muscu
26	92.8	2.2	184364	2 AC128425	AC128425 Rattus no
27	92.6	2.2	150118	2 AC122207	AC122207 Mus muscu
28	92.2	2.2	222585	2 AC101391	AC101391 Mus muscu
29	92	2.2	4995	3 PFAME5A	PFAME5A Plasmodium
30	92	2.2	247175	2 AC129593	AC129593 Mus muscu
31	91.8	2.2	158255	2 AC112577	AC112577 Rattus no
32	91.2	2.2	631	3 PPRPK51	PPRPK51 X53026 P.falciptaru
33	91	2.2	2069	6 E10125	E10125 DNA encodin
34	91	2.2	3399	6 E10126	E10126 DNA encodin
35	91	2.2	207750	2 AC109224	AC109224 Mus muscu
36	90.8	2.2	120044	2 AC129042	AC129042 Rattus no
37	90.6	2.2	153477	2 AC006278	AC006278 Plasmodiu
38	90.6	2.2	170032	2 AC116673	AC116673 Mus muscu
39	90.4	2.2	128921	2 AC112671	AC112671 Mus muscu
40	90.4	2.2	131310	2 AC114893	AC114893 Felis cat
41	90.2	2.2	81019	8 AB073159	AB073159 Arabidops
42	90	2.2	131310	2 AC114893	AC114893 Felis cat
43	90	2.2	143716	2 AL627248	AL627248 Dantis rex
44	90	2.2	198520	2 AC129001	AC129001 Rattus no
45	90	2.2	205221	2 AC115723	AC115723 Mus muscu

ALIGNMENTS

RESULT 1

AX489468

LOCUS AX489468

DEFINITION Sequence 6768 from Patent WO02053728.

ACCESSION AX489468

VERSION AX489468.1 GI:22323480

KEYWORDS

SOURCE

ORGANISM

Candida albicans.

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE

1 Roemer, T., Jiang, B., Boone, C., Bussey, H., and Ohlsen, K. L.

Gene disruption methodologies for drug target discovery

Patent: WO 02053728-A 6768 11-JUL-2002;

Patent: 1794 bp DNA linear PAT 16-AUG-2002

Elutra Pharmaceuticals, Inc. (US)
 Location/Qualifiers
 Source 1.1794
 /organism="Candida albicans"
 /db_xref="taxon:5476"
 BASE COUNT 706 A 286 C 322 G 480 T
 ORIGIN

Query Match 42.9%; Score 1767.2; DB 6; Length 1794;
 Best Local Similarity 99.4%; Pred. No. 7.9e-288;
 Matches 1784; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

134 ATGATTAATGCCAAGCAGACACATTAATCAAAATATAGAAAGATGAAACAAAGCA 193
 1 ATGATTAATGCCAAGCAGACACATTAATCAAAATATAGAAAGATGAAACAAAGCA 60
 194 TACAAATCAATATGTAATAGTACTGATATTCAGATGATCAAAATATAGATCA--TATT 251
 61 TACAAATCAATATGTAATAGTACTGATATTCAGATGATCAAAATATAGATCAATATT 120
 252 TCAGAAAGAAAACTGCCATGCTGAAGATGCCATCGATTATGAAAGATATAGCAAGATTG 311
 121 TCAGAAAGAAAACTGCCATGCTGAAGATGCCATCGATTATGAAAGATATAGCAAGATTG 180
 312 GCAGAAAGAAAGATGATGATGAGATCTACCCAGGAGTGAAGCAATTAATGATTAAC 371
 181 GCAGAAAGAAAGATGATGATGAGATCTACCCAGGAGTGAAGCAATTAATGATTAAC 240
 372 TCAGAAAGAAAGATGATGATGAGATCTACCCAGGAGTGAAGCAATTAATGATTAAC 431
 241 TCAGAAAGAAAGATGATGATGAGATCTACCCAGGAGTGAAGCAATTAATGATTAAC 300
 432 CTTCAGAAAGGAGGAGCTGTAATGCAAAATGATGAAGAAATGCGAGCTCAAGCTGCTGCT 491
 301 CTTCAGAAAGGAGGAGCTGTAATGCAAAATGATGAAGAAATGCGAGCTCAAGCTGCTGCT 360
 492 GAATCTCAATTTGATGCTTTGTTTGAAACTCTAATGATTTGATGATTAATTTAGCCAC 551
 361 GAATCTCAATTTGATGCTTTGTTTGAAACTCTAATGATTTGATGATTAATTTAGCCAC 420
 552 CATGATCATATGCGGCGGAGACATGATGATTAATGATGATTAATGATTAATGATTAATG 611
 421 CATGATCATATGCGGCGGAGACATGATGATTAATGATGATTAATGATTAATGATTAATG 480
 612 AACGACACGATGTTTATTCACATTTAGGAAATGATTAATGATTAATGATTAATGATTAATG 671
 481 AACGACACGATGTTTATTCACATTTAGGAAATGATTAATGATTAATGATTAATGATTAATG 540
 672 AATGATGCTTTGATGATTTAGTGAATCTATTGATGATCAACAAGAGAGACATGATTT 731
 541 AATGATGCTTTGATGATTTAGTGAATCTATTGATGATCAACAAGAGAGACATGATTT 600
 732 ATCAACATCAAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 791
 601 ATCAACATCAAGAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 792 GAAGATCAAAAAAGAAAGAAATTAACGACATTTGAAAGGCAAAAAATCTACAAAAGATT 851
 661 GAAGATCAAAAAAGAAAGAAATTAACGACATTTGAAAGGCAAAAAATCTACAAAAGATT 720
 852 GTTAAACATCTTGAAAGAAAGAAATTCAGACAAATTAATTAATTTCCCTACTTAT 911
 721 GTTAAACATCTTGAAAGAAAGAAATTCAGACAAATTAATTAATTTCCCTACTTAT 780
 912 TCAGACATAGACCATTAATTTTCCACAAATTTTTCACAAAGTCTCAGTATTCAGCT 971
 781 TCAGACATAGACCATTAATTTTCCACAAATTTTTCACAAAGTCTCAGTATTCAGCT 840
 972 TATCAAAAGACAGCAATGCTTGTCGAAAAATATATTAACCTTAACCTAACAAAGTA 1031
 841 TATCAAAAGACAGCAATGCTTGTCGAAAAATATATTAACCTTAACCTAACAAAGTA 900
 1032 AATCTTGAATTCGAGGTTGATCAAAAGAAATTTTCAATTAAGAAAGTCTGATCTGCA 1091

901 AATCTTGAATTCGAGGTTGATCAAAAGAAATTTTCAATTAAGAAAGTCTGATCTGCA 960
 1092 TCGTTGTACACGAGAACAAATATGTCACCAATATTACTCAAGATGCTTGATTTATC 1151
 961 TCGTTGTACACGAGAACAAATATGTCACCAATATTACTCAAGATGCTTGATTTATC 1020
 1152 AAAAATTTAGAAAGCAAAAGATCTTCTATTGATGCTGTTTATTAAGAAATTTAGCGTT 1211
 1021 AAAAATTTAGAAAGCAAAAGATCTTCTATTGATGCTGTTTATTAAGAAATTTAGCGTT 1080
 1212 AAACGATTTGACCTAATATGCGACAAAGTTTGATCTATTATTCGAAAGATTTAGTCTATCT 1271
 1081 AAACGATTTGACCTAATATGCGACAAAGTTTGATCTATTATTCGAAAGATTTAGTCTATCT 1140
 1272 ACCGCTGATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1331
 1141 ACCGCTGATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
 1332 AAGCAATCAATGATGCTTTTGTCTCAACAATCCCTTGACAAATGATTAAGAAATTAAGCA 1391
 1201 AAGCAATCAATGATGCTTTTGTCTCAACAATCCCTTGACAAATGATTAAGAAATTAAGCA 1260
 1392 AAAATCGAATGACCAATAGCTACCAACAATTAACCAAAACCAATAGTAATGTCGAAGAT 1451
 1261 AAAATCGAATGACCAATAGCTACCAACAATTAACCAAAACCAATAGTAATGTCGAAGAT 1320
 1452 GAGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1511
 1321 GAGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 1512 ATGATGATCTTAACTTGTATTTGTTGCTAGTAAGAAAGATGATGATGATGATGATGATG 1571
 1381 ATGATGATCTTAACTTGTATTTGTTGCTAGTAAGAAAGATGATGATGATGATGATGATG 1440
 1572 GTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1631
 1441 GTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 1632 GAATTTATGAAAGAAAGATTAACCAACCAAAACCAAAAGATCTCAATGATGATGATGATG 1691
 1501 GAATTTATGAAAGAAAGATTAACCAACCAAAACCAAAAGATCTCAATGATGATGATGATG 1560
 1692 GAACATTCAGTTCCCGCATGCGATTAACAGACCTTATTAAGTCAAACTTAGCACA 1751
 1561 GAACATTCAGTTCCCGCATGCGATTAACAGACCTTATTAAGTCAAACTTAGCACA 1620
 1752 GATGAAACGAGATCATTCATCAGACCAAGTGTTAATGTCAGACCTGATACATTGGTAGC 1811
 1621 GATGAAACGAGATCATTCATCAGACCAAGTGTTAATGTCAGACCTGATACATTGGTAGC 1680
 1812 TTTTCTTAATTTGAAGTTGCGGAAGCGGAAAAAGACAAAGGGAATCTTTGCAACAGATT 1871
 1681 TTTTCTTAATTTGAAGTTGCGGAAGCGGAAAAAGACAAAGGGAATCTTTGCAACAGATT 1740
 1872 TTTTCCAAAACCTAGGACCTTGACAGTGTGATGATGATGATGATGATGATGATGATGATG 1925
 1741 TCCTCCAAATCTAGTGAATCTTGACAGATGCTGATGATGATGATGATGATGATGATGATG 1794

RESULT 2
 SCYGR274C/C SCYGR274C 3950 bp DNA 1linear PLN 11-AUG-1997
 LOCUS DEFINITION S.cerevisiae chromosome VII reading frame ORF YGR274C.
 ACCESSION Z73059.1
 VERSION Z73059.1 GI:1323498
 KEYWORDS
 SOURCE
 ORGANISM
 Saccharomyces cerevisiae.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE
 1 (bases 1 to 3950)
 Panzeri, L., Agostoni Carbone, M.L., Melchiorretto, P., Plevani, P.,

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 09:41:49 ; Search time 5373 Seconds
(without alignments)
12406.593 Million cell updates/sec

Title: US-09-601-965C-1
Perfect score: 4116
Sequence: 1 aaaaaaaaaaacagctccccc.....tattatgcattctattgtc 4116

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estrov:*
- 6: em_estrpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gp_est1:*
- 10: gp_est2:*
- 11: gp_hic:*
- 12: gp_est3:*
- 13: gp_est4:*
- 14: gp_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gp_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227.8	5.5	907	17	CNS06URB
2	219.8	5.3	940	17	CNS075F8
3	212	5.2	753	17	CNS06VZL
4	199	4.8	984	17	CNS06GB9
5	116.8	2.8	891	17	AZ683582
6	113.8	2.8	518	17	AZ928433

7	108.2	2.6	912	17	AZ551092
8	107.2	2.6	877	17	AZ531291
9	105.6	2.6	816	17	AZ535744
10	104	2.5	997	17	CNS005TE
11	101.2	2.5	908	17	AZ548467
12	99.8	2.4	906	17	AZ551618
13	99.8	2.4	906	17	BH153606
14	98.6	2.4	900	17	BH153980
15	96.2	2.3	976	17	BH149983
16	95.4	2.3	580	13	BJ433319
17	94.8	2.3	574	13	BJ433734
18	94.8	2.3	931	13	BH160272
19	94.6	2.3	849	17	AZ546009
20	94.4	2.3	823	17	AZ676218
21	94.2	2.3	887	14	BQ731479
22	92.8	2.3	843	17	BH139532
23	92.4	2.2	577	13	BJ428840
24	92	2.2	890	17	BH146886
25	91.4	2.2	905	17	AZ550256
26	91	2.2	483	13	CNS02CV8
27	91	2.2	934	13	BM415088
28	90.8	2.2	795	17	AZ528485
29	90.2	2.2	838	17	CNS03W93
30	90	2.2	1135	17	CNS033CQ
31	89.6	2.2	880	17	AZ529191
32	88.8	2.2	589	13	BJ408820
33	88.4	2.1	630	13	BJ443268
34	88.4	2.1	705	17	AQ051650
35	88.4	2.1	739	17	AZ441705
36	88.2	2.1	966	13	BM415686
37	88	2.1	709	13	BJ390371
38	87.2	2.1	461	9	AU262882
39	87	2.1	527	17	AZ279446
40	86.6	2.1	592	17	AZ015337
41	86.6	2.1	694	17	AG044360
42	86.6	2.1	1036	17	CNS03LMJ
43	86.4	2.1	948	13	BM415494
44	86	2.1	1223	17	B12981
45	85.8	2.1	495	17	AZ461180

ALIGNMENTS

RESULT 1
CNS06URB
LOCUS 907 bp DNA linear GSS 06-JUL-2001
DEFINITION T7 end of clone AX0AA012E10 of library AX0AA from strain CBS 7064
ACCESSION AL116205
VERSION AL116205.1 GI:12196372
KEYWORDS GSS.
SOURCE Pichia farinosa.
ORGANISM Pichia farinosa.
REFERENCE Souciet J.L., Aigle M., Artiguenave F., Blandin G., Bojotin-Pukhara M., Bon E., Brotier P., Casaregola S., de-Montigny D., Dujon B., Durrens P., Lepingle A., Llorente B., Maupertuy A., Neveuglise C., Ozier-Kalogeropoulos O., Potier S., Saurin M., Tekala F., Toffano-Nicoche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
AUTHORS Yeast species for molecular evolution studies
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 907)
AUTHORS de Montigny J., Spehner C., Souciet J., Tekala F., Dujon B., Wincker P., Artiguenave F. and Potier S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 09:51:39 ; Search time 134 Seconds
(without alignments)
9420.012 Million cell updates/sec

Title: US-09-601-965C-1
Perfect score: 4116
Sequence: 1 aaaaaaaaaaagctccccc.....tattatgcatcttattgtc 4116

Scoring table:
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/6C_COMB.seq:*
6: /cgn2_6/prodata/1/ina/6D_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.2	2.8	7218	1	US-08-232-463-14
2	75.2	1.8	696	4	US-09-461-697-193
3	75.2	1.8	699	4	US-09-461-697-191
4	75.2	1.8	717	4	US-09-461-697-189
5	75.2	1.8	774	4	US-09-461-697-187
6	75.2	1.8	819	4	US-09-461-697-185
7	75.2	1.8	1669	4	US-09-461-697-184
8	68.4	1.7	72604	4	US-09-268-992-7
9	68.4	1.7	72604	4	US-09-268-992-7
10	67.8	1.6	2223	4	US-08-257-073-4
11	64.6	1.6	44453	4	US-09-146-053-5
12	64.6	1.6	45566	4	US-09-146-053-6
13	63.8	1.6	5394	3	US-08-688-376-1
14	63.2	1.5	1559	4	US-09-019-095A-7
15	62.6	1.5	240	4	US-08-628-417-6
16	62.6	1.5	15231	3	US-09-128-155-16
17	62.6	1.5	176373	3	US-09-128-155-17
18	61.6	1.5	3211	2	US-08-574-859A-8
19	61.6	1.5	3211	2	US-08-574-859A-8
20	61.6	1.5	3901	2	US-08-574-859A-6
21	61.6	1.5	3901	2	US-08-574-859A-6
22	61.4	1.5	289	4	US-09-007-005-17
23	61.4	1.5	289	4	US-09-244-796-17
24	60.6	1.5	376	2	US-08-623-906A-18
25	60.4	1.5	291	1	US-07-922-723A-7
26	60.4	1.5	291	1	US-07-922-723A-7
27	60.4	1.5	291	1	US-08-074-275-7

28	60.4	1.5	291	1	US-08-480-366-7	Sequence 7, Appli
29	60.4	1.5	291	2	US-07-952-277A-7	Sequence 7, Appli
30	60.4	1.5	494	2	US-08-332-766A-22	Sequence 22, Appli
31	60.2	1.5	2169	4	US-09-434-408-3	Sequence 3, Appli
32	60.2	1.5	5935	4	US-09-178-973B-17	Sequence 17, Appli
33	60.2	1.5	5935	4	US-09-418-568F-29	Sequence 29, Appli
34	60.2	1.5	5935	4	US-09-354-243B-29	Sequence 29, Appli
35	59.8	1.5	2791	4	US-09-570-367C-1	Sequence 1, Appli
36	59.4	1.4	2447	2	US-09-014-969-14	Sequence 14, Appli
37	59	1.4	1236	4	US-08-741-134-5	Sequence 5, Appli
38	59	1.4	5361	4	US-08-973-462-2	Sequence 2, Appli
39	59	1.4	6152	4	US-08-973-462-1	Sequence 1, Appli
40	58.4	1.4	333	4	US-09-018-584A-27	Sequence 27, Appli
41	57.8	1.4	361	4	US-09-018-584A-9	Sequence 9, Appli
42	57.8	1.4	1298	3	US-08-948-705-3	Sequence 3, Appli
43	57.8	1.4	1447	4	US-09-443-041A-27	Sequence 27, Appli
44	57.6	1.4	43795	3	US-08-742-185-101	Sequence 101, App
45	57.4	1.4	282	4	US-09-461-697-205	Sequence 205, App

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 768
REFERENCE/DOCKET NUMBER: 30472/114 IMNU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)816-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZcpT-F1s
US-08-232-463-14
Query Match 2.8%, Score 114.2, DB 1, Length 7218;

```
Best Local Similarity 11.6%; Pred. No. 2.2e-15;
Matches 59; Conservative 271; Mismatches 179; Indels 0; Gaps 0;

QY 2332 AAATGTAGTTTCAGAGTATGAATGCTTAGAGTCCAGTATTTCTGTGAAGATG 2391
Db 1576 AATATTTAAATATGAAGTGTGATTTGACTAGCATCGCTTCTAGAGCATCTTA 1517
QY 2392 TTTCACAACTTTCTGTAGCATATGCAATATGCAAAAGATGAAGAATTTA 2451
Db 1516 TTTCAGTTTCAAAAACGGCATGTATGCTCACTGTAATTAATCTATGCAAGTAGTTA 1457
QY 2452 TGGAAATACCAAGACAGTATGAGACCAAGGTATTTGAAAGATGAAGATTAAGCATG 2511
Db 1456 AAGAGATAGAAAGATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1397
QY 2512 TTAATTCAGAGAAAGAAATTTGAACTATGATTAACACCAAGATTTCTCATTTGATG 2571
Db 1396 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1337
QY 2572 ACACATTCATTTGACAGCAGATCTAGATGATATATATGTTTATTCGCTGACAGCA 2631
Db 1336 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1277
QY 2632 GCAGACAGATCATCTAGATTAATAAAGGTATTAAGAGAGATTCATTCGCGATG 2691
Db 1276 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1217
QY 2692 ATGCAAGAAATGAGATATATTAATTAAGATTAAGAGAGAGTTGAGAAAGAAAG 2751
Db 1216 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1157
QY 2752 AACAGAGAAAGAGAAAGAAAGATTAAGAGAAAGACAGAGCAAGAGAAAGAAAG 2811
Db 1156 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1097
QY 2812 ACAAAACCGAAGAGAAATCAAGAA 2840
Db 1096 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1068

RESULT 2
US-09-461-697-193
; Sequence 193, Application US/09461697
; Patient No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasuri
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-193

Query Match 1.8%; Score 75.2; DB 4; Length 696;
Best Local Similarity 49.1%; Pred. No. 2.3e-07;
Matches 231; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

QY 2442 AAAGATTTATGAAATACCAAGACAGGATGAGACCAAGGATTTGAAAGTAAAGAGA 2501
Db 130 AAAGAAATGTCTGAAAGTAAAGAAACAAATATTTGAAAGATGACACAGAAAGAGAGAGA 189
```

```
QY 2502 TTAACGATGTAATTCACAGGAAG-----AGAAATTCGAATCTAGATTACACAGAA 2555
Db 190 AAGAAAGAACAGTGGACAGCAAGTAATAAATAAGAAAGAGATCAGAAAGAGATGAA 249
QY 2556 GATTCCTCATTTGATGACACATGCAATTTGACAGCAAGTATAGATGTAATATGTTG 2615
Db 250 GAAGATCAAAACGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
QY 2616 TTAATCGGTGACAGAGACCAAGATCATCTAGAAAGTAAAGAGAGAGAGAGAGAG 2675
Db 310 GAATATGAAAGAGAGATTAATAATGAAATGAAAGAGAGAGAGAGAGAGAGAGAGAG 369
QY 2676 GATTCATTCGCGATGATGACGAAATATGAGATGATATTAATTAAGATTAAGAGAGAG 2735
Db 370 GATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
QY 2736 GTTGAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2795
Db 430 GATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
QY 2796 AAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2855
Db 490 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 549
QY 2856 ACTGAATTTGATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2905
Db 550 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599

RESULT 3
US-09-461-697-191
; Sequence 191, Application US/09461697
; Patient No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasuri
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-191

Query Match 1.8%; Score 75.2; DB 4; Length 699;
Best Local Similarity 49.1%; Pred. No. 2.4e-07;
Matches 231; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

QY 2442 AAAGATTTATGAAATACCAAGACAGGATGAGACCAAGGATTTGAAAGTAAAGAGA 2501
Db 133 AAAGAAATTTGTGAAAGTAAAGAAAGAAATATTTGAAAGTGCACAGAAAGAGAGAGAA 192
QY 2502 TTAACGATGTAATTCACAGGAAG-----AGAAATTCGAATCTAGATTACACAGAA 2555
Db 193 AAGAAAGAACAGTGGACAGCAAGTAAATAAGAAAGAGATCAGAAAGAGATGAA 252
QY 2556 GATTCCTCATTTGATGACACATGCAATTTGACAGCAAGTATAGATGTAATATGTTG 2615
Db 253 GAAGATCAAAACGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 312
QY 2616 TTAATCGGTGACAGAGACCAAGATCATCTAGAAAGTAAAGAGAGAGAGAGAGAG 2675
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 20:53:18 ; Search time 46 seconds
(without alignments)
3360.233 Million cell updates/sec

Title: US-09-601-965c-2

Perfect score: 6072

Sequence: 1 MEDLRDEALINGLNNNNNN.....CPLYKKWVLGIDDSAAVVG 1160

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6072	100.0	1160	20	AAV30128
2	1578	26.0	1066	19	AAW33632
3	1569	25.8	1066	17	AAW91300
4	765.5	12.6	1872	17	AAW06078
5	765.5	12.6	1872	18	AAW25030
6	765.5	12.6	1886	22	ABG06009
7	761.5	12.5	1575	23	AAW84357
8	760	12.5	1872	15	AAW56493
9	756	12.5	1893	15	AAW56491
10	756	12.5	1893	17	AAW06082

11	756	12.5	1893	18	AAW25020
12	756	12.5	1924	22	ABG06008
13	731.5	12.0	2065	22	ABW66356
14	693	11.4	1234	20	AAV30129
15	297	4.9	740	16	AAW68838
16	296	4.9	740	13	AAW27530
17	293	4.8	1193	21	AAW18306
18	286.5	4.7	1192	21	AAW18165
19	285.5	4.7	2485	21	AAW18172
20	283	4.7	1979	21	AAW18171
21	279	4.6	5024	22	AAW82935
22	278.5	4.6	10182	22	ABW38314
23	275.5	4.5	1812	22	ABW58022
24	269.5	4.4	1166	21	AAW18268
25	269	4.4	2206	21	AAW18254
26	263.5	4.3	1516	21	AAW18195
27	261.5	4.3	1780	22	AAW38681
28	259.5	4.3	1788	22	AAW40467
29	256.5	4.2	665	21	AAW18278
30	255.5	4.2	980	21	AAW18294
31	254	4.2	2285	20	AAW98149
32	254	4.2	6815	22	ABW66811
33	248	4.1	2688	22	AAW40883
34	246.5	4.1	4134	20	AAV31946
35	245	4.0	2663	22	AAW39097
36	244	4.0	1247	21	AAW18215
37	239.5	3.9	1052	22	AAW60276
38	237	3.9	2633	22	ABW06505
39	234	3.9	3696	23	ABP40235
40	233.5	3.8	2295	17	AAW18180
41	232.5	3.8	1089	17	AAW01896
42	232.5	3.8	1089	20	AAV05835
43	232.5	3.8	1089	21	AAV98055
44	231.5	3.8	2573	21	AAW18234
45	231	3.8	1351	21	AAW18290

ALIGNMENTS

RESULT 1
ID AAV30128 standard; Protein: 1160 AA.
AC AAV30128;
XX
XX 14-OCT-1999 (first entry)
XX
XX A TAF-145 protein of Candida albicans.
DE
XX
XX TAF-145 protein of Candida albicans.
KW
XX fungal gene transcription; fungal infection.
XX
XX
OS
XX
XX Candida albicans.
XX
XX
XX W09940199-A2.
XX
XX
XX 12-AUG-1999.
XX
XX
XX 08-FEB-1999; 99WO-US02940.
XX
XX
XX 09-FEB-1998; 98US-0024100.
XX
XX
XX (SCRT-) SCRTPTGEN PHARM INC.
XX
XX Long F, Thompson CM, Wobbe RC;
XX
XX WPI: 1999-494297/41.
XX
XX N-PSDB: AAX86690.
XX
XX Novel fungal protein critical for expression of fungal genes
XX
XX provides a new target for antifungal drugs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:27:24 ; Search time 32 Seconds
(without alignments)
3484.873 Million cell updates/sec

Title: US-09-601-965c-2

Perfect score: 6072

Sequence: 1 MEDLPKDEALINGLNNSNNNN.....CPYKKKVLGIDDDSAAVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*

1: pir1:*\n2: pir2:*\n3: pir3:*\n4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1578	26.0	1066	2	S50237	TATA box-binding p
2	1537	25.3	979	2	S62469	probable transcrip
3	768	12.6	1865	1	I48155	transcription init
4	756	12.5	1893	1	A40262	transcription init
5	727	12.0	2068	2	A47371	transcription init
6	607	10.0	1490	2	S32373	DNA-binding protei
7	508.5	8.4	1994	2	D86452	protein F6N18.13 l
8	325	5.4	3394	2	T18501	hypothetical prote
9	297.5	4.9	1738	2	T14867	interactin - slime
10	294	4.8	3724	2	T18427	hypothetical prote
11	293.5	4.8	743	2	A29232	101k malaria antiq
12	289	4.8	1193	2	C71605	hypothetical prote
13	286.5	4.7	1192	2	A71623	probable secreted
14	285.5	4.7	2339	2	A45597	DNA-directed RNA p
15	285.5	4.7	2437	2	T18482	hypothetical prote
16	285.5	4.7	2485	1	H71621	serine/threonine-s
17	283	4.7	1979	2	C71622	hypothetical prote
18	282	4.6	4550	2	T18440	hypothetical prote
19	275	4.5	1526	2	A45605	mature-parasite-in
20	275	4.5	1790	2	S67593	transport protein
21	269.5	4.4	1166	2	H71609	hypothetical prote
22	269	4.4	2139	2	T18296	myosin heavy chain
23	269	4.4	2206	2	G71611	hypothetical prote
24	265	4.4	1780	2	T17272	hypothetical prote
25	264.5	4.4	1650	2	T18444	hypothetical prote
26	263.5	4.3	1127	2	T28317	ORF MSV156 hypothe
27	263.5	4.3	1516	2	E71619	RAD2 endonuclease
28	262	4.3	1871	2	D96796	probable heat choc
29	261.5	4.3	1312	1	BMBYDL	Rad50 protein - ye

30	259	4.3	2401	2	T28676	thoperty protein -
31	258.5	4.3	1422	2	T18404	chromatin remodell
32	258.5	4.3	2910	2	T28156	DNA-directed RNA p
33	256.5	4.2	665	2	B71609	hypothetical prote
34	256	4.2	1658	2	S55101	hypothetical prote
35	255.5	4.2	980	2	E71606	hypothetical prote
36	254	4.2	2285	2	T12796	probable transglyc
37	250	4.1	1313	2	A48467	myosin heavy chain
38	250	4.1	1805	1	A64224	hypothetical prote
39	248.5	4.1	2022	2	T43214	ovtl protein - nem
40	248	4.1	2526	2	T20531	hypothetical prote
41	248	4.1	2722	2	T20532	hypothetical prote
42	248	4.1	2738	2	E88320	protein F07A11.6 l
43	246.5	4.1	1939	2	T18372	repeat organellar
44	246	4.1	2663	1	S28261	centromere protein
45	244.5	4.0	1254	2	T18277	kinesin heavy chai

ALIGNMENTS

RESULT 1
S50237
TATA box-binding protein-associated factor chain TAF1145 - yeast (Saccharomyces cere
N:Alternate names: protein G9374; protein YGR274c; TAF1130 protein
C:Species: Saccharomyces cerevisiae
C:Date: 27-Jan-1995 #sequence-revision 10-Feb-1995 #text-change 21-Jul-2000
C:Accession: S50237; S62389; S62397; S62394; S62395; S64607
R:Reese, J.C.; Apone, L.; Walker, S.S.; Griffin, L.A.; Green, M.R.
submitted to the EMBL Data Library, September 1994
A:Description: Yeast TAF115 in a multisubunit complex required for activated transcri
A:Reference number: S50237
A:Accession: S50237
A:Molecule type: DNA
A:Residues: 1-1066 <RE>
A:Cross-references: EMBL:U14954; NID:9595837; PID:9595838
R:Reese, J.C.; Apone, L.; Walker, S.S.; Griffin, L.A.; Green, M.R.
Nature 371, 523-527, 1994
A:Title: Yeast TAF115 in a multisubunit complex required for activated transcriptio
A:Reference number: S62389; MUID:95021683; PMID:7935765
A:Accession: S62389
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1066 <RE>
A:Cross-references: EMBL:U14954; NID:9595837; PID:AA79178.1; PID:9595838
A:Accession: S62397
A:Molecule type: protein
A:Residues: 583-599; 651-671 <REF>
R:Poon, D.; Bai, Y.; Campbell, A.M.; Bjorklund, S.; Kim, Y.J.; Zhou, S.; Kornberg, R.
Proc. Natl. Acad. Sci. U.S.A. 92, 8224-8228, 1995
A:Title: Identification and characterization of a TFIID-like multiprotein complex fro
A:Reference number: S62390; MUID:95396770; PMID:7667272
A:Accession: S62394
A:Molecule type: DNA
A:Residues: 370-830 <POO>
A:Accession: S62395
A:Molecule type: protein
A:Residues: 368-384; 528-554; 752-783 <POW>
R:Panerai, L.; Agostoni Carbone, M.L.; Melchiorretto, P.; Plevani, P.; Martegani, E.;
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64591
A:Accession: S64607
A:Molecule type: DNA
A:Residues: 1-1066 <PAN>
A:Cross-references: EMBL:273059; NID:91323498; PID:e243586; PID:91323499; MIPS:YGR274
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:TAF145; TAF130
A:Cross-references: SGD:S0003506; MIPS:YGR274c
A:Map position: 7R
Query Match
Best local Similarity 26.0%; Score 1578; DB 2; Length 1066;
32.4%; Pred. No. 1.4e-61;

Matches 391: Conservative 204: Mismatches 353: Indels 260: Gaps 37:

```

OY 38 GQPELTDEMAAQAASQDALFG-----NSNDPDSNT-----SHHDH 77
DB 8 GKPLNLADEDE-----AYBAIFEGEGSLEIGSYIGDEGANSKDYETHLPDAVFEDEDE 62
OY 78 MGSNGSIIDNNHSSVNDHD-----GLEPNLGNHLLDDDDNDGN-----DLGEL 124
DB 63 LADDDDLPEE--SDANLHPAMTMTGAADVNGAVLIDNSINMLPRLINGDSOQ 119
OY 125 F---D-----OQEDSNVIN--TKKHLD-----DPSNN 148
DB 120 FILEDDGTPATSNALFPMGMANEIHLATEGVLDGSGANEIGHSLQSLIGVNGNDMSIN 179
OY 149 DGKTAOEDQEKENKROKROKLOKIVKHEKQIKRNTIKYPTYSRHRPFNK--FFS 207
DB 180 GGFIMEDMSGKKHKATKLD-----LINH-EKYLKK-----YPPDEKGIILKKNKLIIR 230
OY 208 PSPQYKIQAPALSKNIKPLIPKVNLEIVDOKKIKRLSADTASLSHEDKNVTNIT 267
DB 231 RSVYHMHSE-----ISRYKKRPWMLKFKVOODDKRLNSRITSYVAPLYQGN--NLL 284
OY 268 QDDLOFINKLESKRSSIDSFIKEIDYKRDWNCDFH---YSKDLVLTDDDDAI 323
DB 285 QSN-----SSASRKLIVSIDSLEPPIKQOK--RKIIHDEKTSIDLLATDMDQEKI 338
OY 324 IMAGDNEYSIVKPINELLNPNLNSKONROKIENTNTNNYNNNSNVODEED----- 378
DB 339 INQSTSTATL-----ADSSMTPNLKF-----SGGY-KLKSLIEDVADMOWDE 381
OY 379 DDJFNGOINLDR-LKLDNDPMLLEVPK-----KYDATKSVVPSYDKELEKFI 428
DB 382 DMIDAKLEKSKHMLNDEKLLMIETNNMLAQOKQODSSNLLPLMETTIOQKFNLI 441
OY 429 SNODEVELLRKNYNTKOSQSLNLSNIHNSVPLRLQTPYKVLSTDETSFHRPV--N 486
DB 442 SNDDKYOILKTHQKVVASTISNLNIHNSOPALNDSPTKYAVNPKQLRHFRENCSH 501
OY 487 VRPSTLVSFKLKRKKDKKSLQOIFSKTSIDLTADTGNLIALEYSQYRILSNFG 546
DB 502 IRPRTKIVFSKLARKKRRDKGDKVSKSTSDLTIGDAPYLLMEYSQTPALSKFG 561
OY 547 MGSLLINYKKEPRNDISPRKAQIGETHILGVEDNSPEMNFEVAPGDEVPYLLNNVRA 606
DB 562 MANKLINYYRRANEDQTLRPLRVGETHVLGVDKSPFNWEGFVERGHVPTLYNNMIRA 621
OY 607 PIRKHNKPTPELLVKSQAGSHQKFFYLGINFNFAVNGTFPV--EYVAPHSKVTNISK 665
DB 622 PVFNHDSIGTDFLLTKSSGFGISNRFYLRNINHLFTVGOTFVPEEIRGPNRSKVTSMKAT 681
OY 666 RLKVVNVRVMSLSGPRISVQKVSXKHPRESDMONRORLEKFEYEROGEDQWYKRG 725
DB 682 RLKMIIRYILNHNISKAISIDIRIAKHFPODYGONKQVKEFKYORDGPEKGLMRKD- 740
OY 726 NDVPRGEIEITITPESSLMDTMOFGQOVLDDNVLNVLFGESRQSSRSRKKDKRDSI 785
DB 741 DEKLDDNAVASLITPEGISQVESMSQGLQFQEDN----- 775
OY 786 ADDAENGDITKKDEKEVEKEKEREKREKDKKDEKDKREKKEKSKSKEQDEI 845
DB 776 --EAYND-----SKLVS----- 786
OY 846 DVEELAPWNLNRNFVIANQTKMLQNGEDPTGIGLFSMLRATQKNEFKPLTPRPE 905
DB 787 -LEENLLPNTTKKEINSTOMRAMTQIHGVDPGTGEGFSFLTKSMKGGFVKSPPSN 845
OY 906 NVPR-----SNAAHNOKLYEOEIKRIWYSORSSLVDHCEGESKLOQIYNEYPAD 957
DB 846 NNSNKKGTNHSYNAOQOKAYDEEIAKTYTHKS----- 882
OY 958 HELYLRKLEDDOQVOQOQOQDSLOADOQOQOQOQONRVLRITRRVRDENGIVHAKVEFI 1017
DB 883 --LISNPF-----EMNPNDEINOTNNKHVKTDRDDKILKIVRKRKRENGIIGROCIIFI 935

```

Query Match 25.3% Score 1537: DB 2: Length 979:
 Best local similarity 31.7% Pred No 7.7e-60
 Matches 374: Conservative 196: Mismatches 349: Indels 262: Gaps 31:

```

OY 1018 HDBRLTRAVYKRRKQIEDELLKNADVDEILPTNDEKLNIRKRLAEEKLANLEKRAKQSR 1077
DB 936 RDBRVIOGYIKIKEDQKEDVKNKLEEDTSKTNLEEEK--OKKLOLELANLEKSOQRA 994
OY 1078 AKRPKDLHAAAGATTIDANTVMLPDGSYVIGKGI--GKRSRPRCKNCGAYGHR 1136
DB 995 ARONSK-----FNGATRTENSVDNCSDLAGYDGAARNGKRNTRCATCGOIGHIR 1048
OY 1137 TNAKCPY 1144
DB 1049 TNKSCPMY 1056

```

RESULT 2
 Probable transcription initiation factor tlfid subunit - fission yeast (Schizosacchar
 C:Species: Schizosaccharomyces pombe
 C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000
 C:Accession: J36580; 562469
 R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 Submitted to the EMBL Data Library, October 1995
 A:Reference number: Z21745
 A:Accession: J36580
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-979 <BA2>
 A:Cross-references: EMBL:Z54354; NID:G1019398; PIDN:CAA91179.1; PID:G1019412; GSPDB:G
 A:Experimental source: strain 972h-; cosmid c2g11
 A:Gene: SPAC2G11.14
 A:Map position: 1L
 A:Introns: 21/3; 212/1; 261/3
 C:Keywords: transcription initiation

Query Match 25.3% Score 1537: DB 2: Length 979:
 Best local similarity 31.7% Pred No 7.7e-60
 Matches 374: Conservative 196: Mismatches 349: Indels 262: Gaps 31:

```

OY 9 AINGLNSNNNNHDKDDDDDEFNRLLQEGP-----ELTDEMAAQAASQDALFG 63
DB 2 SFQGLIVENENTKSGYNDGNDLTLDFKONGTMSVINSLLDTPNNGMESKILDSFE 61
OY 64 NSNDEDS--NISHHHMGDSNGIIDDNNHSSVNDHDLFNNLGNLHLLDDDDNDGLD 122
DB 62 NSNPQDGPVYEDDFMG-----SIHKEFGNNINEMDD---MEDVS 98
OY 123 ELFDQOEDSNVINTFKKHLDDSDNNDKTAQEDQEKENKROK----- 167
DB 99 D--DNLPRERQAVNVTGDKDDEDE--GKLAKEMGEAAQOVLSGVFSIPSGLVPSSE 153
OY 168 -KOKLOKIVKHEKE--QIKRNI--KYRPTYSRHRPFNKFFSPQYRYRORPITALSK 224
DB 154 PKSTVSTTEBEIQNEAQIRRSIVKTFEPFERGVLLNPSLEFKRP-----VKLAPPKK 208
OY 225 NIKPLIPKVNLEIVDOKKIFKRLSADTASLSHEDKNVTNITQDDDLDFINKLESKR 284
DB 209 TRKVCYVGRITLEVVDYALITFSK-----KSLPLKRVV 243
OY 285 DSFIKEIDYKRDWNCDFDHYSKDLVLTDDDDAIINAGDNEYSIVKPINELLN 344
DB 244 SPISTHKRRRTANTSORNDGLDNLNFTTNDKEKNTIDESP----- 287
OY 345 PLDNSKONKQKIENTNTNNYNNNSNVQ-----DEEEDDINGOINLDRKLDNDMP 398
DB 288 -----VNKTNQSFFIDKSLVDIDAFAFENIRFDGDTGTSKVVNLNDP 330
OY 399 NLIFVPS--KYDATKSVVSTDK--LLELFENISNDEVELLRKNYNTKOSQSLNLE 455
DB 331 KLLDQPLRKEDSQRSPADTHQRNSLAWKFNISNDPAVEYMLKONHOSKVRNTLSQALIE 390
OY 456 HSNVPLRLQTPYKVLSTDETSFHRPVNVRGIVTSFKLIRKRRKDKGKSLQOIF 515

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:01:08 ; Search time 21 Seconds
(without alignments) 2291.073 Million cell updates/sec

Title: US-09-601-965c-2

Sequence: 1 MEDLPDEAINGLNSNNNN.....CPLYKKMVLGIDDSAAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1578	26.0	1066	T145_YEAST	P06813 saccharomyc
2	1537	25.3	979	T111_SCHPO	Q09813 schizosacch
3	765.5	12.6	1872	T2D1_HUMAN	P21675 homo sapien
4	723	11.9	2068	T2D1_DROME	P51123 drosophila
5	293.5	4.8	743	ABRA_PLAFC	P22620 plasmodium
6	285.5	4.7	2339	RPCL_PLAFA	P27620 plasmodium
7	274.5	4.5	1790	USO1_YEAST	P23386 saccharomyc
8	261.5	4.3	1312	RA50_YEAST	P12753 saccharomyc
9	256	4.2	1658	YMG7_YEAST	Q03661 saccharomyc
10	250	4.1	1805	HMM2_MYCCE	P47460 mycoplasma
11	248.5	4.1	2022	ANT1_ONCYO	P21249 onchocerca
12	246	4.1	2663	CENE_HUMAN	Q02224 homo sapien
13	242.5	4.0	1875	MUP1_YEAST	Q02455 saccharomyc
14	235	3.9	1332	SPY7_YEAST	P35177 saccharomyc
15	233.5	3.8	1938	MYSD_CAEEL	P02567 caenorhabd
16	233.5	3.8	2869	RBP1_PLAFA	Q00798 plasmodium
17	233	3.8	678	GARP_PLAFA	P13816 plasmodium
18	232.5	3.8	1089	NMD2_YEAST	P38798 saccharomyc
19	227.5	3.7	2245	MYSD_DICDI	P54697 dictyostell
20	225	3.7	2230	GOG4_HUMAN	Q13439 homo sapien
21	224	3.7	1928	MYSL_YEAST	P08964 saccharomyc
22	223	3.7	2349	TFR_HUMAN	P12270 homo sapien
23	222	3.7	1233	YF16_YEAST	P43597 saccharomyc
24	222	3.7	1957	YD86_SCHPO	Q10411 schizosacch
25	220.5	3.6	1664	YIO9_YEAST	P53705 candida alb
26	220.5	3.6	1679	YIO9_YEAST	P40457 saccharomyc
27	219	3.6	1480	IFN1_YEAST	P39520 saccharomyc
28	218.5	3.6	1245	YUN1_YEAST	P53935 saccharomyc
29	217.5	3.6	899	MYH3_YEAST	Q04500 saccharomyc
30	217.5	3.6	3210	MYH3_CHICK	P10587 gallus gall
31	217.5	3.6	3210	CENF_HUMAN	P49454 homo sapien
32	216.5	3.6	1898	TRHY_HUMAN	Q07283 homo sapien
33	216.5	3.6	1938	MYA_AEOIR	P24733 aequipecten

34	216.5	3.6	1976	1	MYHA_BOVIN	Q27991 bos taurus
35	214	3.5	2871	1	DESP_HUMAN	P15924 homo sapien
36	211.5	3.5	1324	1	CUT3_SCHPO	P41004 schizosacch
37	211.5	3.5	1630	1	MSPI_PLAFA	P04932 plasmodium
38	211.5	3.5	1639	1	MSPI_PLAFA	P04933 plasmodium
39	211.5	3.5	1818	1	HMM2_MYCPN	P73471 mycoplasma
40	211.5	3.5	2116	1	MYSD_DICDI	P08799 dictyostell
41	211	3.5	1962	1	MYSA_DROME	P05661 drosophila
42	210.5	3.5	1972	1	MYHA_MOUSE	Q08638 mus musculu
43	209.5	3.5	1513	1	STU1_YEAST	P38198 saccharomyc
44	209.5	3.5	1976	1	MYHA_RAT	Q91150 rattus norv
45	207	3.4	1178	1	NNNA_YEAST	P36044 saccharomyc

ALIGNMENTS

```

RESULT 1
T145_YEAST          STANDARD:      PRT: 1066 AA.
ID T145_YEAST
AC P46677;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 35, Last annotation update)
DE Transcription initiation factor TFIID 145 kDa subunit (TBP-associated
DE factor 145 kDa) (TAFFI-145) (TAFFI-130).
GN TAFFI145 OR YGR274C OR G9374.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP RP STRAIN=x57;
RC MEDLINE=95021683; PubMed=7935765;
RX Reese J.C., Apone L., Walker S.S., Griffin L.A., Green M.R.;
RT "Yeast TAFIIS in a multisubunit complex required for activated
RT transcription."
RL Nature 371:523-527(1994).
RN [2]
RP RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97279232; PubMed=9133740;
RA Ruzzi M., Marconi A., Sallola M., Fabiani L., Montebove F.,
RA Frontali L.;
RT "The sequence of a 8 kb segment on the right arm of yeast chromosome
RT VII identifies four new open reading frames and the genes for
RT YTAFFI145,
RT YTAFFI145,
RN [3]
RP RP SEQUENCE OF 368-384; 528-554 AND 752-783, AND CHARACTERIZATION.
RC STRAIN=YPH252;
RX MEDLINE=95399670; PubMed=7667272;
RA Poon D., Bai Y., Campbell A.M., Bjorklund S., Kim Y.-J., Zhou S.,
RA Kornberg R.D., Well P.A.;
RT "Identification and characterization of a TFIID-like multiprotein
RT complex from Saccharomyces cerevisiae."
RT Proc. Natl. Acad. Sci. U.S.A. 92:8224-8228(1995).
RL CC -I- FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID
(IID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF
RNA POLYMERASE TRANSCRIPTION.
CC -I- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs) WHOSE MW RANGE FROM
CC 25-150 KDa.
CC -I- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -I- SIMILARITY: TO S.POMBE TAFII-11. SOME, TO HUMAN TAFII-250 (CCG1)
CC AND TO DROSOPHILA TAFII-230.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:26:54 ; Search time 50 Seconds
(without alignments)
4780.293 Million cell updates/sec

Title: US-09-601-965C-2
Perfect score: 6072
Sequence: 1 MEDPDRDEAINGLNNNNN.....CPYKKVGLGIDDSAAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	12.6	1865	11 060544	060544 mesocricetu
2	760.5	12.5	883	5 08SR44	08SR44 encephalito
3	731.5	12.0	2065	5 097068	097068 drosophila
4	607	10.0	1490	5 09TX96	09TX96 drosophila
5	561.5	9.2	1792	5 09XUL9	09XUL9 caenorhabdi
6	536.5	8.8	1810	10 09LJ62	09LJ62 arabidopsis
7	508.5	8.4	1994	10 09LPI9	09LPI9 arabidopsis
8	399.5	5.4	1946	5 09J291	09J291 plasmodium
9	328	5.4	3130	5 09BK46	09BK46 plasmodium
10	325	5.4	3394	5 077384	077384 plasmodium
11	308	5.1	2867	5 09N2M3	09N2M3 plasmodium
12	300	4.9	959	5 08TIP2	08TIP2 dictyosteli
13	298	4.9	440	11 035361	035361 mus musculu
14	287.5	4.9	1387	5 09G276	09G276 plasmodium
15	287.5	4.9	1738	5 076329	076329 dictyosteli
16	295.5	4.9	3254	5 09BK45	09BK45 plasmodium

17	294	4.8	3724	5 077320	077320 plasmodium
18	291.5	4.8	2081	10 09LH98	09LH98 arabidopsis
19	291	4.8	1461	5 08ST04	08ST04 dictyosteli
20	290.5	4.8	791	5 08T207	08T207 dictyosteli
21	289	4.8	1193	5 096257	096257 plasmodium
22	286.5	4.7	1114	5 097242	097242 plasmodium
23	285.5	4.7	1192	5 096127	096127 plasmodium
24	285.5	4.7	2423	5 077393	077393 plasmodium
25	285.5	4.7	2485	5 096134	096134 dictyosteli
26	285	4.7	1699	5 095PH3	095PH3 dictyosteli
27	283	4.7	1979	5 096133	096133 plasmodium
28	282	4.6	4550	5 077336	077336 plasmodium
29	280.5	4.6	739	5 09N1H1	09N1H1 plasmodium
30	278	4.6	1661	5 006166	006166 plasmodium
31	275.5	4.5	1812	5 09VZD9	09VZD9 drosophila
32	275	4.5	1510	5 025920	025920 plasmodium
33	275	4.5	1790	3 007380	007380 plasmodium
34	271.5	4.5	2771	5 026216	026216 plasmodium
35	270	4.4	3628	5 0968Y0	0968Y0 plasmodium
36	269.5	4.4	1166	5 096219	096219 entamoeba
37	269	4.4	2139	5 007569	007569 plasmodium
38	269	4.4	2206	5 096205	096205 plasmodium
39	268.5	4.4	16215	5 09NFS3	09NFS3 drosophila
40	265	4.4	1780	4 09DFR5	09DFR5 homo sapien
41	265	4.4	2038	5 0967Y0	0967Y0 dictyosteli
42	264.5	4.4	1650	5 077328	077328 plasmodium
43	263.5	4.3	1127	12 09YVT6	09YVT6 melanoplus
44	263.5	4.3	1516	5 096154	096154 plasmodium
45	263.5	4.3	1583	4 015045	015045 homo sapien

ALIGNMENTS

RESULT 1

ID 060544 PRELIMINARY: PRT: 1865 AA.

AC 060544;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CCG1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94215915; PubMed=8163200;
RA Hayashida T., Sekiguchi T., Noguchi E., Sunamoto H., Ohba T.,
RA Nishimoto T.;
RT "The CCG1/TAIR1250 gene is mutated in thermosensitive GI mutants of
the BHK21 cell line derived from goldenhamster.";
RL Gene 141:267-270(1994).
DR EMBL; D26114; BAA05110.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 2.
DR PRINTS; PR00503; Bromodomain.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS00633; BROMODOMAIN_2; 2.
SQ SEQUENCE 1865 AA; 211866 MW; A81614946CC0CF24 CRC64;

Query Match 12.6% Score 768; DB 11; Length 1865;
Best Local Similarity 22.1%; Pred. No. 5.6e-28;
Matches 306; Conservative 226; Mismatches 445; Indels 408; Gaps 51;

QY 43 TNDDEMAQAAMAEQFDALFGNSND-----PDSNISH----- 75
I : : : : : | | | | |
DB 18 TQSDDESSGGGPPSLTGLFNGINGAGQLGEGSVLDDECKHLHAGALGSLITELTA 77
: : : : : | | | | |
QY 76 -DHMGGDSNGIIDDN-----HHSSVND-----HDLFNNLNGNHLDDDDND 116

```

Db 78 NEELAGDGLVNDGMRISHEDAVDSDINEVADESRRYQWFGSLQPLCHSADYDED 137
QY 117 GLNDLCELEF-----DOOEDSNVINTKHKHLLDDSDNDKTOEDQKCKEN 162
Db 138 DYDADDEDIDCKLMPRPPIGVYKKEKDDGLTGEKVDSSSSDSESEMGPOEQAQAES 197
QY 163 KROLKROKLOKIVKHEKEQIKRNIKYYFPYSHRPFNFHFKFSPS---POYRYQCPRA 219
Db 198 KQKLTPLPLAGIMQH-DATKILPSVTELPFEFRPGKVLRLRLGPGKNVSVRSARRK 256
QY 220 IALSKNIPPLIPKVV-----NLEIEVDQKKIFK-----LRSA 251
Db 257 -RKKKHREPIOEBOIEEESVELEVQKSLMNYDAPRPPREOCLSDDELTMAPEBSK 315
QY 252 DVALSHEDK-----NVTINQDDDLFIKNLESKSSIDSFKEI 291
Db 316 FQOSTGDTDKVMDTKPRVAEKRYGPARLWYDMLGVEDGSGFDYGFMRKRETERPAIKCK 375
QY 292 DYVK-RDWTNCKKFDHYSD--LVLTSDWDDAIIINAGNEYSTVKP-----INEL 340
Db 376 MMTKLRLKEESNGIDDLADENFLMTQLHEDDIWDGEDVKHKGTKPRASLAGWLPSS 435
QY 341 LLNPLDLSKONKOKIENDNTNNV-----QNNSNVODEEEDDIFNGQ-----INLKL 391
Db 436 MTRNMAVIVQGGFAATLDDDKRWYSIFPIDMEDLYVGHMEDNIWDQAQMPRIEPPVL 495
QY 392 KLDMDPMLF-VPSKKVDATKSVVPSTDK-----LL-----E 423
Db 496 TLDPIDENLLEIPEKEEATSNPSPEKESSLKSRILLGKTGVIKKEEQOQNSQPE 555
QY 424 LK--FNISNDEYELLKRNNTKQ--RSOLSNLNEHSVAPLRLQRYKVKVKTSTDETRS 479
Db 556 VKDFPMNSNDEY-----YKQOGLRGTFGGNITONHSIPAVELRPPFRTHNGPILKRG 609
QY 480 FHRPVF-----NVRPCTLVSFSLKLRKKKDKGSLQOQFESTK-SQLTVA 524
Db 610 FHRPRLKYSFGALSQPCPHSVOPRLKHKKKAKMRQORQASGGEMFMTPODLTKG 669
QY 525 DTGNIIALEVSEQYRPLISNFMGSKLINYKERRNDTSRKAQDGER---HILGEDR 581
Db 670 D-GDLIALEYSEENGRLPMQVGMATIKINYKRRKQPGARDCKGELVYGH-----T 722
QY 582 SPFNFEVAVAGDEVPTLYNNMVARPIGKHONKRPDELLVSSQAGSHQKGYLRGINENF 641
Db 723 SPF--LGSILHPGQLQAFENNI,FRAPITLHKMPETDELITIR---QGYIRLELVDF 775
QY 642 AVGNTFPV-EVPAPHSKRVNTNISKNRKLMVFRVW-NSLGPV-RISVKDVSKNREHSDM 698
Db 776 VVGQOQCLFEVPGNSKRANTHIDFLQVFLYRLKWSKDRPRIRIMEDIKAFPSHSES 835
QY 699 ONKORLKEPEYXOROGDQGWYKGLNDVYIGEEIRMTTTPEDSSLMDTMOFGQOYLD 758
Db 836 STRKRLKLCADFRTGDSMMWVLKS--DFRLPTBEEIRAMVSPQCCAYUUMIAEQLK 894
QY 759 DNNVLFEGOSROESSNRKDKREDSDADDAENGDDINKKEKEVEKEEQEREKEKGD 818
Db 895 D--AGGEGS-----FPAPBEENEDF----- 914
QY 819 KEKDKDKDKTEKEKSKSKEJDETDIVEEELAPMNL,SRNFVANTQTKMLOLNGECDD 878
Db 915 -----QMKIDDEVARTAPMNTTAFIAMGKCLLETGVADP 951
QY 879 TGIQGLFSMLRATQKNPFRPLFPPPENVPKS-----NAAAHNOK--LYEO 922
Db 952 TCGCGEFSYVKIPNK-PTQOKDKKEPQYKTYTGTDADLRSLKLNKAKOQLKRFVGP 1010
QY 923 EIKRLIWSQSSSLVD-----HGECTESKLOQIYNEYPRADH-ELY----- 961
Db 1011 EIKRL-----SRMEVIDVYRTMSTYQARSGBPMKFRAR-GRSFVAHEQERYKEECORIFD 1066
QY 962 LKNKLEQOQV-----QOQOQDPSIQADQOQOQOQOQONR 995

```

```

Db 1067 LONKVLSTFEVLSTDTDSSAEDSDSEEMCKNIENMLONKTSQSLSREPEREKKELOR 1126
QY 996 V-----LRTRRVDENGIYHRRKVEFIHDPILI 1023
Db 1127 MLAAASASAGNNHRDDDTASVTSLSNSTATGRCLKIYRTFREDEGKEYVRCETVRKPAVI 1186
QY 1024 RAYVK-RKQIODELKNADVDEILPTNDKELNKRKALEKLANEKRAKOSRARKP 1082
Db 1187 DAYVIRITTKDEEFIRKFALEDE---OHREMKERRR-IQEDLRRLKRNQEKRLKGP 1242
QY 1083 KDLIHMAAAGATIIDANTVMLPDGSYVJGKGIGKSR---TRCKNGCAGHITNA 1139
Db 1243 E-----KPKKKKRPDLKLKGCAGGALGHMTRNK 1272
QY 1140 KCPV 1144
Db 1273 FCPV 1277

RESULT 2
ID 08SR44 PRELIMINARY: PRY: 883 AA.
AC 08SR44:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Transcription Initiation factor TFIID 11kDa subunit.
GN EC010_0760.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; Pubmed=11719806;
RA Kalinka M.D., Duprat S., Cornillot E., Mettenier G., Thamarat F.,
RA Prenster G., Barbe V., Peyretailade E., Brotier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weisenbach J., Vivas C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001);
RL EMBL: AL590449; CAD25793.1;
SO SEQUENCE 883 AA; 102000 MW; 799568726E6FA5CD CRC64;

Query Match
Best Local Similarity 12.5%; Score 760.5; DB 5; Length 883;
Matches 253; Conservative 149; Mismatches 304; Indels 319; Gaps 35;

QY 150 GRTAODEKKEKKNKROKLOKIVKHEKEQIKRNIKYYFPYSHRPFNFHFKFSPS 209
Db 10 GRAODEESVSSEAGDGPGRSVWRIRFLDLVIEREYKLRKKYGRK----- 58
QY 210 POYRYQCPRALSKNIKPLIPTKVMLEIEVDQKKIFK-----LRSDAFA--SLSHEDKN 262
Db 59 -----VPE-SMSYEVDQAFRSMYVAROSNTLQALSGSTRS 97
QY 263 VTNITQDDDLFIKNLE---SKRSSIDSEFIKIDYKRDWTMCDFHYSKLVYSTTD 317
Db 98 -TGLRQ---YLKRLQLEKEISKRAKALDSDSLREVEVK-----YFDLY---E 137
QY 318 WDDDAIINAGDNEYSIKVINLNNPLDNSKONKOKIENDNTTNNYONNNSVODEEB 377
Db 138 WE-----NNIYDPRDAGGRKMKDITVEF---VDSILLEDW 171
QY 378 DDDIFNGOINLND---LKLMDNPNLLFVPSKKVDATKSVV-----STDKLEEK 425
Db 172 ERYVAVDESDKSKRSFLLTYLEDNLLF---EKIDDKQARPKRKQKEVFSVDKPLK 228

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2003, 21:26:03 : Search time 22 Seconds
(without alignments)
1551.390 Million cell updates/sec

Title: US-09-601-965C-2

Perfect score: 6072
Sequence: 1 MEDLPDEALINGLNNNNNN.....CPLYKKVGLIDDSAAVVG 1160

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1569	25.8	1066	2 US-08-308-818-1	Sequence 1, Appl
2	765.5	12.6	1872	1 US-08-188-582-14	Sequence 14, Appl
3	765.5	12.6	1872	1 US-08-646-715-14	Sequence 14, Appl
4	756	12.5	1893	1 US-08-188-582-11	Sequence 11, Appl
5	756	12.5	1893	1 US-08-646-715-11	Sequence 11, Appl
6	722	11.9	1377	2 US-08-308-818-4	Sequence 4, Appl
7	297	4.9	740	1 US-08-257-073-5	Sequence 5, Appl
8	278.5	4.6	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
9	254	4.2	2285	4 US-09-308-375-2	Sequence 2, Appl
10	234	3.9	3696	4 US-09-134-001C-5080	Sequence 5080, Ap
11	232.5	3.8	1089	1 US-08-375-300-2	Sequence 2, Appl
12	232.5	3.8	1089	1 US-09-177-431-2	Sequence 2, Appl
13	232.5	3.8	1089	5 PCT-US95-16930-2	Sequence 2, Appl
14	230.5	3.6	1664	1 US-09-599-652-2	Sequence 2, Appl
15	230.5	3.6	1664	1 US-08-642-846-2	Sequence 2, Appl
16	220.5	3.6	1664	4 US-09-264-604-2	Sequence 1, Appl
17	220.5	3.6	3248	1 US-08-353-700-1	Sequence 1, Appl
18	220.5	3.6	3248	5 PCT-US95-16216-1	Sequence 4820, Ap
19	219.5	3.6	1211	4 US-09-134-001C-4820	Sequence 28, Appl
20	219	3.6	1211	4 US-08-431-080-28	Sequence 28, Appl
21	219	3.6	1085	4 US-08-938-534-28	Sequence 28, Appl
22	219	3.6	1085	4 US-09-345-294-28	Sequence 94, Appl
23	216.5	3.6	1898	1 US-08-056-200-94	Sequence 94, Appl
24	216.5	3.6	1898	2 US-08-800-644-94	Sequence 6, Appl
25	215.5	3.5	2482	1 US-08-328-254-6	Sequence 29, Appl
26	212.5	3.5	1312	4 US-09-345-882-29	Sequence 4, Appl
27	206.5	3.4	764	1 US-08-375-300-4	Sequence 4, Appl

28	206.5	3.4	764	3 US-09-177-431-4	Sequence 4, Appl
29	206.5	3.4	764	5 PCT-US95-16930-2	Sequence 4, Appl
30	203.5	3.4	1162	2 US-08-728-323A-2	Sequence 2, Appl
31	203.5	3.4	1162	4 US-09-298-568-2	Sequence 2, Appl
32	203.5	3.4	1588	5 PCT-US93-07261-11	Sequence 11, Appl
33	203.5	3.4	1663	5 PCT-US93-07261-16	Sequence 16, Appl
34	203	3.3	1507	3 US-08-929-329-5	Sequence 1, Appl
35	202	3.3	1939	4 US-09-310-187A-1	Sequence 4, Appl
36	197	3.2	1177	4 US-09-134-001C-5106	Sequence 4, Appl
37	196.5	3.2	1435	2 US-08-568-459A-4	Sequence 4, Appl
38	196.5	3.2	1435	2 US-08-487-826B-4	Sequence 148, Appl
39	196.5	3.2	1435	4 US-09-210-288-4	Sequence 51, Appl
40	193	3.2	1312	2 US-08-592-126-148	Sequence 51, Appl
41	193	3.2	1312	2 US-08-687-080-51	Sequence 51, Appl
42	193	3.2	1354	3 US-08-685-871-2	Sequence 5314, Ap
43	190.5	3.1	930	4 US-09-134-001C-5314	Sequence 3, Appl
44	190	3.1	1886	4 US-08-938-105-3	Sequence 4318, Ap
45	188.5	3.1	676	4 US-09-134-001C-4318	

ALIGNMENTS

RESULT 1
US-08-308-818-1
Sequence 1, Application US/08308818
Patent No. 5847077
GENERAL INFORMATION:
APPLICANT: Green, Michael R
TITLE OF INVENTION: A No. 5847077e1 Fungal Multisubunit Protein
TITLE OF INVENTION: Complex Critical for Expression of Fungal Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,818
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0342/0A404
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-52707700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1066 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE: S. cerevisiae
ORGANISM: S. cerevisiae
IMMEDIATE SOURCE:
CLONE: TAF-145
US-08-308-818-1
Query Match 25.8%; Score 1569; DB 2; Length 1066;
Best local similarity 32.3%; Pred. No. 6.6e-104;

Wed Mar 12 10:13:17 2003

us-09-601-965c-2.raii

Page 2

	Matches	390	Conservative	204	Mismatches	354	Indels	260	Gaps
QY	38	GQPLTIDEEMAAQAAASQ	FDALFG	-----	NSNPDSNI	-----	SHDH	77	
Db	8	GKTNLANADE	-----	AVEALIRGEGGSL	ELISYIGCGDCGANSKDT	ETHLPAVD	FEDEDE	62	
QY	78	MGGSGNIIIDNNHHSVNDH	-----	GLFNILG	CNHLDDDNGLN	-----	DLGEL	124	
Db	63	LADDDDLPEE	-----	SDANLHPAMT	MGAYDDV	ENGAVL	IGDSLSL	NOPEIN	DLSSQ 119
QY	125	F---D	-----	QOELSVIN	-----	TYKKHKD	-----	DDSN	148
Db	120	FILEDDGCPATSNALF	MGMLANEI	HLAT	ETGVL	DGSGANEI	GSOLSIG	CVGN	CNDMSIN 179
QY	149	DGKTAOEDQEK	ENKRLQK	QKLOK	LYKH	LEKEDIK	ENIKRYE	PTYS	SHRPFNRK 207
Db	180	GGFIMPEP	MDGKH	KATK	TL	-----	LNH	EKYLK	-----
QY	208	PSFOYRYG	PPALAL	SKNIF	ELI	PTVNL	ELVEDOK	IFIL	RSADTASL
Db	231	RSVPYHMSE	-----	ISRVK	PEP	PLMLK	EVQDD	KRLF	SRITSIYAP
QY	268	QDDLP	EIKL	MSR	SSIDS	IFKID	YV	KRM	TCDFH
Db	285	QSN	-----	SSASR	GLI	HSID	ELP	EIK	QOKR
QY	324	INAGNE	YIVK	PIN	EL	LN	PL	NS	KORUK
Db	339	INOG	TS	ST	ATL	-----	ADSS	MT	NKLF
QY	379	DI	FGQ	I	L	D	M	D	P
Db	382	DM	I	D	A	K	E	S	H
QY	429	SND	VE	L	L	R	K	K	Y
Db	442	SND	Q	Y	O	L	K	T	H
QY	487	V	R	P	T	L	V	F	S
Db	502	I	R	P	G	T	K	I	F
QY	547	M	G	S	L	I	N	Y	R
Db	562	M	A	N	L	I	N	Y	R
QY	607	P	I	F	H	D	K	P	T
Db	622	P	V	R	H	D	I	S	G
QY	666	R	L	K	N	V	R	V	N
Db	682	R	L	K	M	I	T	I	R
QY	726	N	D	V	I	P	E	E	E
Db	741	D	E	L	I	L	N	E	A
QY	786	A	D	A	E	N	G	D	D
Db	776	-	E	A	N	F	D	-----	SKLS
QY	846	D	V	E	E	E	L	A	P
Db	787	-	L	E	E	N	L	P	N
QY	906	N	V	P	K	-----	S	N	A
Db	846	N	S	S	N	K	K	G	T
QY	958	H	E	L	Y	K	N	K	L
Db	883	-	L	S	I	S	N	P	E

```

QY      1018 HDPFLIAAYRKRRKOIDEILKNNADVDEILTPTNKKELTKIRKLAEKANLERAKOSR   1077
Db       936 RDRPVIGGYIKIQDKEDVKNLKEETSKINNLIELEK-QKLTQLDELANLESQQRA    994
QY      1078 AKRPKRLLIHAAAAGATITIDANTVMPLDPSIYGCGGT-GKGKSRTRCKNGCAYGHIR  1136
Db       995 ARONSK-----RNGGATRETNYSVDNSDLGVYTGDNAARNKGNNTTTRCATCGOIGHIR  1048
QY      1137 TNKPCEPY 1144
Db       1049 TKNPCPMY 1056

RESULT 2
US-08-188-582-14
; Sequence 14, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Rupert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weitzler, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-188-582-14

Query Match          12.6%; Score 765.5; DB 1; Length 1872;
Best Local Similarity 22.5%; Pred. No. 4e-46;
Matches 314; Conservative 218; Mismatches 435; Indels 427; Gaps 54;

QY      43 TNDEMAQAAMAAEFSDPDLFGSNSD-----FDNSTSHH-----              75
Db     24 TDSDSDSAGGGCFSLAGLGIMINAGOLEGESVLDDDECKRLLAGLGALGLISLTYTELA   83
QY      76 -DHMGCSNGJTDQN-----HHSSVND--HD-----GLFNNTLGNQHLLDD   113

```